

Cost and time damping: evidence from aggregate rail direct demand models

Andrew Daly^{1,2} · Nobuhiro Sanko³ · Mark Wardman^{1,4}

Published online: 3 June 2016

© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract There is a significant body of evidence from both disaggregate choice modelling literature and practical travel demand forecasting that the responsiveness to cost and possibly to time diminishes with journey length. This has, in Britain at least, been termed ‘Cost Damping’, and is recognised in guidance issued by the UK Department for Transport. However, the consistency of the effect across modes and data types has not been established. Cost damping, if it exists, affects both the forecasting of demand and our understanding of behaviour. This paper aims to investigate the evidence for cost and time damping in rail demand using aggregate rail ticket sales data. The rail ticket sales data in Britain has, for many years, formed the basis of analysis of a wide range of impacts of rail demand. It records the number of tickets sold between station pairs, and it is generally felt to provide a reasonably accurate reflection of travel demand. However, the consistency of these direct demand models with choice modelling and highway demand model structures has not been investigated. Rail direct demand models estimated by ticket sales data indicate only slight variation in the fare elasticity with distance, as is evidenced in the largest meta-analysis of price elasticities conducted to date (Wardman in *J Transp Econ Policy* 48(3):367–384, 2014). This study of UK elasticities shows strong variation between urban and inter-urban trips, presumably a segmentation at least in part by purpose, but less remaining variation by trip length. A lack of variation by length supports the hypothesis of cost damping, because constant cost sensitivity would imply that fare elasticity would increase strongly with distance, because of the increasing impact of higher fares at longer distances. In this paper we indicate that rail direct demand models have some consistency of behavioural paradigm with utility based choice models used in highway planning. We go on to use rail demand data to estimate time and fare elasticities in the context of various

✉ Andrew Daly
andrew@alogit.com

¹ ITS, University of Leeds, Leeds, UK

² RAND Europe, Cambridge, UK

³ Graduate School of Business Administration, Kobe University, Kobe, Japan

⁴ Systra Ltd., Manchester, UK

cost damping functions. Our empirical contribution is to estimate time elasticities on a basis directly comparable with cost elasticities and to show that the phenomenon of cost damping is strongly present in ticket sales data. This finding implies that cost damping should be included in models intended for multimodal analysis, which may otherwise give incorrect predictions.

Keywords Demand analysis · Rail demand · Model estimation · Cost damping

Introduction

There is a significant body of evidence from both disaggregate choice modelling literature and practical transport modelling that the sensitivity to cost and possibly to time diminishes with journey duration; Daly (2010) and Rich and Mabit (2013) give reviews, indicating that the effect may be caused by heteroskedasticity in the travelling population. This has, in Britain at least, been termed ‘Cost Damping’, and has entered demand analysis and appraisal guidance issued by the Department for Transport (DfT 2014).

If such relationships exist in disaggregate data and in car trip matrices, we would expect to find them in other forms of behavioural response. The paradigm of utility maximisation, for example, applies in principle to all choices and travel behaviour. One source of evidence on behavioural response is rail ticket sales data. In Great Britain there exists a large and accurate record of the number of rail trips made between many pairs of stations and this has been extensively exploited to support a wide range of modelling opportunities. As a result there is extensive evidence on how rail demand is influenced by fares, journey time and economic activity, and to a lesser extent the impact of competition from other modes and other service quality factors (ATOC 2013).

In this paper we examine whether there is evidence to confirm such cost damping effects in rail ticket sales data covering a very wide range of flows and many years. The existing results indicate that there is limited variation with distance in the elasticity of demand with respect to fare (‘fare elasticity’). Simply, we would expect that fare elasticity *should* become stronger (more strongly negative) with distance, because longer trips have generally higher fares, so that (in a linear model) a proportional increase on all fares has more impact at longer distances. So a finding of limited variation with distance gives evidence in support of the hypothesis of cost damping. That is, the response mechanism in the model is not linear but shows a decreasing proportional sensitivity to fares as fares increase. In this paper we present further analyses of this issue and compare this with the disaggregate evidence.

The importance of these findings would be to show the consistency of behaviour across different modes and data types, so that consistent utility functions can be used. Practically, the findings indicate how fare elasticity varies with trip length, suggesting ways in which fares could be set to optimise revenue. For modelling, the results indicate that cost damping should be included in models of travel demand.

Additionally, we take the opportunity here to conduct analogous analysis of time elasticities. In addition, conducting these two forms of analysis also provides evidence on how the value of time (VOT) varies with journey distance. VOT is an important parameter in transport planning and one about which there is also much evidence from the disaggregate literature, but little evidence derived from aggregate data.

The analysis here is based upon tens of thousands of observations. The basic approach is to estimate fixed-effect pooled cross-section time-series models but instead of the standard approach of estimating constant elasticities with respect to fare and time we allow the fare and time elasticities to vary with journey length in ways suggested by disaggregate and conventional transport planning analyses. This is done by specifying continuous functions which allow an increasing, diminishing or no effect from journey length, measured by distance, cost or journey time, on the elasticities, as empirically justified.

The following section of the paper discusses the formulation of the models to be used for analysis and how a near-equivalence can be obtained between the models of conventional transport demand analysis and those previously used for rail ticket sales data. Section 3 describes the data that is used and presents the results of analyses of this data using the various model forms proposed. Section 4 presents the summary results and gives a discussion and conclusions.

Formulation of hypotheses

While the data on which the main analysis of this paper is based comprises substantial numbers of observations, each of these observations contains relatively little detail about the travellers. Thus, analysis of this data yields very solid results with limited insight. The question to which the paper is then addressed is whether insights from other areas of travel analysis, in particular discrete choice, can be transferred to the estimation context of large-scale ticket sales data. We focus particularly on Cost Damping.

Cost Damping

Cost Damping is the feature in some travel demand models whereby the sensitivity of the model to marginal changes in time and/or cost declines as journey lengths increase. It has been observed in a wide range of choice modelling contexts (for a review and discussion of potential causes, see Daly and Carrasco 2009) and is incorporated in many models used for practical transport planning in the United Kingdom. Recent changes to official Guidance (DfT 2014) have accepted that it may be necessary to include such variation in practical models of travel demand.

The DfT Guidance draws from a detailed study of Cost Damping in theory and practice. The report of that study (Daly 2010) found that effectively all of the practical urban regional, and national travel demand models that were in use in the UK were based on generalised cost functions; many used some form of Cost Damping. Daly (2010) suggested that cost or time budgets were not relevant to the phenomenon, but that heteroskedasticity in preferences is a likely cause. Four distinct types of Cost Damping were found to be used in practice to adapt the generalised cost functions, as illustrated in the following table adapted from the report.

The first classification, represented by the columns of the table, indicates whether the mechanism operates on the entire generalised cost function, or separately on the components of the function, e.g. the time and cost. The second classification distinguishes between transformations in which the variation over trip length is fixed, most often by relating it to the travel distance, and those in which it is a function of the travel variables only, most often a non-linear function of time and cost. The report advises against the use

of distance in such functions, as there is no behavioural basis for such use, but the practice is very convenient and therefore widespread.

The models used in practice, whether based on local estimation using discrete choice modelling or on the transfer of aggregate mode and destination choice models with local adjustment, are all of the tree-nested logit form, in which the utility U of the elementary alternatives is specified (for a given demand segment and alternative) as

$$U = V + \varepsilon = \lambda G + B + \varepsilon$$

where V is the ‘representative’ utility; ε is an error term assumed to follow an appropriate ‘GEV’ distribution; G is the generalised ‘cost’¹ function

$$G = t + c/v$$

where t is the travel time; c is the travel cost; v is the value of time; $\lambda < 0$ is the model sensitivity, which needs to be provided, e.g. by estimation from data; B is a constant applicable to the specific alternative.

With the GEV assumption for the distribution of the error term, this framework implies the prediction of demand by a logit-type model, based on the representative utilities V .

Particularly for public transport applications, the term t above should be understood as ‘generalised journey time’ comprised of weighted components such as access time, waiting time and ‘line-haul’ time; to emphasise this point we use the notation GJT for this variable. A correction may be applied to account for crowding. Other variables may also be included in the generalised cost function to account for interchanges and other service aspects.

The term B may account for destination-specific as well as mode-specific aspects of choice. A normalisation will usually be required because it is not possible to identify a full set of such constants.

The term λ is implied to be constant across modes in the formulation above and indeed this constraint is often applied in practice. However, this practice may be criticised as it does not take account of different levels of comfort etc. that apply to the different modes. In the present paper, we are concerned purely with train journeys and the issue is not of concern.

In the simplest form of these logit models, a function linear in time and cost components (as shown above) is applied. Cost damping is the process of introducing non-linear, ‘downward curving’ functions into this process. Specific formulations for each of the cost damping mechanisms in Table 1 are presented and discussed in Sect. 2.3, but first we consider the differences between the modelling approaches of regional travel demand modelling and classical ticket sales data analysis.

Interpretation for ticket sales data analyses

In conventional transport planning approaches based on mode and destination choice, demand is predicted by a tree-nested logit model, for a given origin:

$$T_{mj} = T_0 \frac{\exp V_m}{\sum_{m'} \exp V_{m'}} \cdot \frac{\exp V_{mj}}{\sum_{j'} \exp V_{mj'}} \quad \text{with} \quad V_m = \theta \log \sum_{j'} \exp V_{mj'}$$

$$T_{mj} = T_0 p_m p_{jm} = T_0 p_{mj}$$

where T_0 is the total number of trips generated at the origin; $V_{mj} = \lambda(t_{mj} + c_{mj}/v) + B_{mj}$ is the representative utility for travel by mode m to destination j , including generalised time

¹ Generalised ‘cost’ is conventionally quantified in time units.

Table 1 Classification of cost-damping mechanisms

Operating on	Entire function	Separate components
Fixed function ‘differential scaling’	I	II
Dependent on costs ‘transformed variables’	III	IV

and cost divided by the value of time as described above; V_m is the representative utility for mode m ; θ is a parameter, $0 < \theta \leq 1$, indicating the relative sensitivity of mode and destination choice; p_m is the probability of choosing mode m ; $p_{j|m}$ is the probability of choosing destination j , given that mode m is chosen; $p_{mj} = p_m p_{j|m}$ is the probability of choosing destination j and mode m .

In the formulation presented, which is most common in European models (see e.g. Fox et al. 2003), destination choice is modelled as more sensitive to time and cost than mode choice.

This model can be differentiated and rearranged to give

$$\frac{\partial \log T_{mj}}{\partial V_{mj}} = 1 - (1 - \theta)p_{jm} - \theta p_{mj} \quad (1)$$

In contrast, the typical formulation of the models used for the analysis of rail ticket sales data is, for a given flow segment,

$$\log T = B + V + \varepsilon$$

where B is a route-specific constant; V contains variables in log or linear form describing the service on the route and other relevant variables such as income and ε is an error term.

In the conventional transport planning models, the variables are most often in linear form, though log and intermediate functions are now being used more widely. However, for rail applications the variables are almost always presented in log form. For example, typical rail demand models would be of the following form, for a given flow k and time period t

$$\log T_{kt} = B_k + \sum_r \eta_r x_{rkt} + \sum_s \delta_s x_{skt} + \varepsilon_{kt}$$

where η are coefficients indicating sensitivity with respect to continuous variables; x_r are continuous variables such as generalised time, cost and income, represented in a log form (so that the η s are elasticities); δ measures the impact of dummy variables; x_s are dummy variables applying for various effects; ε is an error term, assumed to follow a distribution consistent with the use of least-squares analysis of the data.

For these rail models we can therefore write the simple equation

$$\frac{\partial \log T_{kt}}{\partial V_{kt}} = 1$$

Comparing this with the Eq. (1) for tree-nested logit models presented above, it is easy to see that the rail ticket sales data models correspond to the logit models if:

- the probabilities $p_{j|m}$ and p_{mj} are so small they can be neglected (or p_{mj} is small and θ is close to 1) and
- we neglect also any difference between the distributions of the error term.

Rail demand from one origin is distributed over many destinations, while rail typically has a small share of the total market from any origin and destination, so it does not seem unreasonable to expect the relevant probabilities to be small. However, for some origins (e.g. in the commuter belt of large cities) these assumptions may be less accurate. The difference between the assumptions concerning the error term are of less concern.

Therefore, it seems reasonable, at least as an approximation, to draw a close parallel between the ‘utility’ function V_{mj} used in conventional planning and the function V_{kt} used in rail data analysis. In particular we may test the cost damping functions derived for disaggregate modelling and conventional transport planning in the context of ticket sales data.

Model formulation for analysis

Formulating the cost-damped utility functions of the previous section for this type of analysis leads to the models shown in Table 2, dropping the subscripts for flow and time period in the interest of clarity. Model 0 represents the basic linear model.

For models of types I to IV, the functional form for the cost damping function needs to be specified. There is a wide range of possibilities and among these we have chosen to work primarily with power functions. For Models of Type III and IV we additionally tested log functions, as set out in Table 2.

In Table 2, d indicates the distance between origin and destination, while λ and α are parameters to be estimated. B denotes effects specific to the origin and destination. t represents the generalised journey time GJT . Certain limits can be imposed on these parameters, for example that we would not easily accept models in which the sensitivity to time or cost increased with increasing trip length, as this is contrary to a large body of experience.

A further restriction on model form is the ‘kilometrage test’ presented by Daly (2010). This test is based on the common-sense notion that if the price of travelling per kilometre is increased one would not expect more kilometres to be travelled. It is not a strict economic argument, but general economic thinking suggests that it is a sensible requirement to impose on a model. In the context of rail data analysis the test should be slightly less

Table 2 Model forms proposed

Type		Constraints
0	$\log T = B + \lambda_1 c + \lambda_2 t$	$\lambda < 0$
I	$\log T = B + (\lambda_1 c + \lambda_2 t) d^{-\alpha}$	$\lambda < 0$ $0 < \alpha < 1$
II	$\log T = B + \lambda_1 c d^{-\alpha_1} + \lambda_2 t d^{-\alpha_2}$	$\lambda < 0$ $0 < \alpha < 1$
III	$\log T = B + \lambda_0 (c + \lambda_2 t)^\alpha$	$\lambda_0 < 0, \lambda_2 > 0$ $0 < \alpha < 1$
IIIA	$\log T = B + \lambda_0 \log (c + \lambda_2 t)$	$\lambda_0 < 0, \lambda_2 > 0$
IV	$\log T = B + \lambda_1 c^{\alpha_1} + \lambda_2 t^{\alpha_2}$	$\lambda < 0$ $0 < \alpha < 1$
IVA	$\log T = B + \lambda_1 \log c + \lambda_2 \log t$	$\lambda < 0$
IVB	$\log T = B + \left(\begin{array}{l} \lambda_1 \alpha_1 c + \lambda_1 q_1 (1 - \alpha_1) \log c \\ + \lambda_2 \alpha_2 t + \lambda_2 q_2 (1 - \alpha_2) \log t \end{array} \right)$	$\lambda < 0$ $0 < \alpha < 1$ $q \text{ constant} > 0$

rigorous, because the connection between price and distance is not at all precise. Nevertheless, it seems reasonable to impose on our models the requirement that an increase in price per distance will not cause the distance travelled to increase. Similarly, a reduced speed should not cause the distance travelled to increase and this test on time can be extended, admittedly with less precision, to the generalised time *GJT*. These tests impose limits on the functions we can use for cost damping.

It must always be the case that demand declines with increases in time and cost and this imposes the sign constraints on λ as shown in Table 2. The table also indicates the range of values of the parameters α that will give cost damping and maintain consistency with the kilometrage test. Normally we would not accept values outside these ranges. Additionally, for models IV and IVB, we should require that $\alpha_1 \leq \alpha_2$, so that VOT does not decrease with trip length. Model IVB interpolates between linear and log form.

For these models, elasticities with respect to time and cost can be derived, as follows.

In most of these models, there is a fixed relationship between the time and cost elasticities based on the concept that sensitivity to time and cost is determined exactly by VOT. In models other than IV, IVA and IVB, the ratio of time to cost elasticity is t^v/c , as can be seen by calculating the ratios of the elasticity columns and comparing with the column giving v , and v does not depend on time or cost, so that a test of independent time and cost elasticity parameters is essentially an estimation of VOT. However, in models IV, the situation is more complicated and VOT varies with the time and cost. Models IV thus indicate the relative strength of time and cost damping (see Table 3).

In Model IVA the elasticity is constant, so that this model corresponds to the classical rail demand analysis.

Some of the model forms proposed include power functions and in these cases, when the optimum value of the power is close to zero, estimation of the model can prove difficult. In these cases, it is sometimes useful to replace the simple power function by the Box-Cox transformation

$$x^{(\alpha)} = \frac{x^\alpha - 1}{\alpha}, \quad \text{for } \alpha \neq 0; \text{ and } x^{(\alpha)} = \log(x) \quad \text{for } \alpha = 0.$$

This function is continuous at $\alpha = 0$, facilitating estimation in some cases. When appropriate, we have made this substitution.

Table 3 Elasticities of proposed model forms

Type	Cost elasticity	Time elasticity	Value of Time v
0	$\lambda_1 c$	$\lambda_2 t$	λ_2/λ_1
I	$\lambda_1 c/d^\alpha$	$\lambda_2 t/d^\alpha$	λ_2/λ_1
II	$\lambda_1 c/d^{\alpha_1}$	$\lambda_2 t/d^{\alpha_2}$	$\lambda_2/\lambda_1 d^{\alpha_1-\alpha_2}$
III	$\alpha \lambda_0 c G^{\alpha-1}$	$\alpha \lambda_0 \lambda_2 t G^{\alpha-1}$	λ_2
IIIA	$\lambda_0 c/G$	$\lambda_0 \lambda_2 t/G$	λ_2
IV	$\lambda_1 \alpha_1 c^{\alpha_1}$	$\lambda_2 \alpha_2 t^{\alpha_2}$	$\lambda_2 \alpha_2 t^{\alpha_2-1}/\lambda_1 \alpha_1 c^{\alpha_1-1}$
IVA	λ_1	λ_2	$\lambda_2 c/\lambda_1 t$
IVB	$\lambda_1 (\alpha_1 c + q_1 (1 - \alpha_1))$	$\lambda_2 (\alpha_2 t + q_2 (1 - \alpha_2))$	$\frac{\lambda_2 c (\alpha_2 t + q_2 (1 - \alpha_2))}{\lambda_1 t (\alpha_1 c + q_1 (1 - \alpha_1))}$

Note that $G = (c + \lambda_2 t)$ the generalised cost of the journey

Results of analyses based on these models are reported in Sect. 3.

The role of income

The sensitivity of demand to the cost of rail tickets depends, of course, on travellers' incomes. Information is available on average incomes by region and by year and this data has been used to adjust the monetary variables entering the model. In every case, incomes and fares are corrected for inflation to the price level of 2005. However, adjustments have also been made for changes in real income levels.

The data used for analysis relates to a number of years, over which incomes have changed, and to a number of regions, over which incomes vary. Income enters the model in two ways.

- First, the level of income affects the overall level of demand; as described below, this is accounted for by estimating a coefficient that links the demand to the income (adjusted for inflation) per capita for each region and year.
- Second, the impact of fare is mitigated by income; that is, regions and years with higher incomes are more willing to accept higher fares. In the modelling, this is accommodated by multiplying the fares by the factor $\overline{income}/income_{rt}$

where $income_{rt}$ is the inflation-adjusted income for the origin region r and year t and \overline{income} is its average over all r and t included in the dataset.

In some cases it has been suggested that an 'income elasticity' should be applied, i.e. applying a power function to the income to adjust the fares. However, more recent work (Abrantes and Wardman 2011; Daly and Fox 2012²) has found that an elasticity of 1 is more consistent with recent data. Moreover, selected tests made on the current data also indicated that unit elasticity was most appropriate. Subsequently, when we refer to fare we mean the fare adjusted by income in the direct way indicated above.

The income measure used in these analyses was the regional Gross Value Added per capita. We believe we are using the best available income data for this modelling, although the average income for an area may not represent the income of the potential rail-using population. Income elasticity is not the main focus of this work, but is an essential point in understanding the generalised journey time (GJT) and fare elasticities which are our main focus.

Analysis results

Data chosen for analysis

Rail ticket sales data in Great Britain has, for many years, formed the basis of analysis of a wide range of impacts on rail demand. It records the number of sales between stations and, with some exceptions such as urban trips in metropolitan areas where the use of area wide travelcards are common, it is generally felt to provide a reasonably accurate reflection of rail demand.

² These studies consider the impact of varying income on the value of time, which effectively converts cost variables to a time scale for use in modelling. Then an income elasticity of 1 for the value of time implies an income elasticity of the sensitivity of behaviour to cost also equal to 1.

To conduct this analysis, a number of data sets were available to us. The data selected relates to trips outside London and not wholly within the South East, and station pairs are limited to those separated by a distance of 20–300 miles. The London and South East trips are excluded since a wide range of tickets has historically been on offer for those journeys, while other flows tend to be on a smaller range of tickets with less competition between them. Hence the use of average revenue per trip as a measure of fare involved fewer approximations. We also excluded station pairs with distances under 20 miles where rail is often not an attractive transport mode and those over 300 miles where fewer trips are observed and air competition is a relevant issue.

By excluding London trips, there is less chance of one destination dominating choice, or of rail becoming the dominant mode. These two features make it easier to accept the approximate equivalence of model form between rail modelling and general transport planning discussed above, where it was important that the choice probabilities for the alternatives should be low.

The remaining data covers 3201 station-to-station movements for the years 1990–2005, excluding 1994 which was seriously affected by widespread industrial action, and all sales other than season tickets. Sales were aggregated for each station-to-station pair and each year.³ Pooling data across routes and over time yields 48,015 observations, each representing a station pair and a year, for modelling purposes.

Rail ticket sales data can make only limited distinctions by journey purpose, so there might be a concern that the cost damping is in part driven by journey purpose variations by distance. Our data covers non-season tickets which, along with the distances involved, means that commuting will be a very minor proportion of trips. Although business travel with its lower price but higher time elasticities might be more prevalent for longer distance journeys, our Non-London flows are, with the exception of a very few movements between key business centres, dominated by trips for other purposes. Whilst specific other purposes might be expected to vary by distance, such as holidays and short breaks forming a larger proportion of longer trips and visiting friends or relatives, entertainment and personal business trips forming a larger proportion of shorter distance trips, we have no evidence that these trip purpose shares or, more importantly, their elasticities vary strongly with distance.

Endogeneity of explanatory variables in econometric models can lead to biased coefficient estimates. The classic railway example is that a high frequency on a route can stimulate high demand but high frequency is often a result of high demand in order to accommodate the passengers. It is the former relationship that concerns us here but without separating the two effects it might be feared that we will obtain an exaggerated effect of the impact of service frequency. We do not though regard endogeneity here to be a particular issue for a number of reasons.

The conventional wisdom in Great Britain is that endogeneity is likely to be much more of an issue in cross-sectional models, as in the example above, rather than in models as estimated here (with fixed effects controlling the cross-sectional effects), where the variation informing the parameter estimates is the time-series rather than cross-sectional dimension. As far as we are aware, not since Jones and Nichols (1983) has endogeneity been raised as an issue in either the academic or practitioner literature. They were addressing concerns about previous research, largely cross-sectional since sufficiently long time-series were then not available, and they concluded that endogeneity was not likely to

³ Return tickets were included both for outbound and for return legs, but attributed to the origin station of the outbound leg as this links better to regional income.

arise in studies with a strong time-series element. Our view is that there are a number of compelling reasons why simultaneity bias is not a material problem in the models reported here.

Firstly, and very importantly, we are here analysing station-to-station data and it is not practical, or indeed customarily attempted (as is confirmed by the relevant industry body), for a railway company to price or offer a particular level of service between two stations that is closely based on the level of demand between those two stations. Fares and to a slightly lesser extent the service offering are largely corridor based, relating to distances travelled to avoid anomalies,⁴ on which there are very many station-to-station movements with very different demand levels. It would be hard to then claim that the price or service level offered on the flows in our regression analysis was strongly demand dependent.

Secondly, some fares in the UK have a historical basis and have been increased (roughly) in line with inflation or are constrained by the competitive situation which is outside the control of the operator. Indeed, some station-to-station movements, particularly in our data set, cover two or more operators. These features again break the endogenous link with demand. Whilst advance tickets are sold on a yield management basis, whereby greater demand leads to higher prices, this phenomenon tends to be more recent than our data set and in any event is much less prevalent on the Non-London based flows investigated here.

Finally, all journeys in Britain have at least one fare whose level is regulated by government. This clearly weakens any endogeneity.

We therefore conclude endogeneity is not a serious problem, and in particular we see no reason why it has distorted the relationships between demand elasticities and distance which, in our models, depend on fare differences rather than absolute fare levels.

Classical rail models

The basic approach is to estimate fixed-effect pooled cross-section time-series models, whereby a constant gives the basic magnitude of demand on each flow, linked to factors such as the population around the origin and destination stations and competition from other modes, with other variables explaining variations around this level of demand. Instead of the classical approach of estimating constant elasticities with respect to fare and time the models allow the fare and time elasticities to vary with journey length in ways suggested by disaggregate and conventional transport planning analyses. As described in the previous section, this is done by specifying continuous functions which allow increasing, diminishing or no effect with journey length, measured by distance, cost or journey time, on the elasticities.

The theoretical formulations discussed in Sect. 2.3 of this paper have to be extended for practical use, giving a general formulation for all the models of

$$\log T_{ij}^t = \text{constant}_{ij} + f(\text{fare}_{ij}^t, GJT_{ij}^t, \text{dist}_{ij}, \text{Inc}_i^t, \lambda, \alpha) + \beta_1 \log \text{Inc}_i^t \\ + \beta_2 \text{Hat2000}^t + \beta_3 \text{Hat2001}^t + \beta_4 \text{Hat2002}^t + \text{error}$$

where T_{ij}^t denotes rail ticket sales for year t and station pair ij ; constant_{ij} indicates the 3201 ‘fixed effects’ station-to-station dummies for each flow ij ; Inc_i^t denotes income per head in year t in the region where origin station i is located, adjusted for inflation as described in

⁴ In some further analysis we found a strong dependence of fares on distance and years, but little or no relationship to demand.

Sect. 2.4; $\text{Hat2000}^t = 1$ for $t = 2000$; 0 otherwise; and Hat2001^t and Hat2002^t are defined similarly.

The function f varies across the model types as described in Sect. 2.3. The variables appearing in this function are the following.

fare_{ij}^t gives the fare between the stations ij in year t , derived as revenue per trip, in pounds sterling defined at 2005/6 price levels and adjusted by income as described in Sect. 2.4.

GJT_{ij}^t gives the generalised journey time which is a measure in time units (minutes one way) of the timetable-related service quality and comprises the origin to destination station journey time, service headway and any need to change trains. The GJT measure is standard in the UK rail industry and was supplied in its combined form.

dist_{ij} represents the distance in miles from i to j .

The estimated parameters α , λ represent parameters related to the damping and parameters expressing the relative importance of fare and GJT.

This formulation includes dummy variables for the year of the Hatfield accident (2000) and for each the next 2 years to represent the disruption to services due to widespread speed restrictions and engineering works on the rail network. The impact of this work was not uniform across the network and it is not to be expected that an impact will be measured significantly, particularly in the simple way we have used, in all the models. Parameters β are estimated for this effect and, more importantly, for log income. The inclusion of these parameters is essential to avoid biasing the estimates of the key time and cost parameters of interest.

Parameters including 3201 values for constant_{ij} are estimated by nonlinear least squares. The code is written in GAUSS (Aptech Systems, Inc. 2015) with some modifications from a code by Hill and Adkins (2001) to handle the large number of fixed effect constants. The least-squares approach is equivalent to a maximum likelihood estimation, providing that the error terms are normally distributed, an assumption that is supported by the use of log trips as the left-side variable.

Rail demand models, at least in Great Britain, are typically estimated in constant elasticity form as in Model IVA. Table 4 reports the fixed effects model of this type, excluding the coefficients for the 3201 route-specific variables.

The result indicates that the data set provides a robust basis for the investigation that is here to be undertaken. Not only are the key parameters estimated with an extremely high level of precision, the elasticities are generally plausible and align reasonably well with the

Table 4 Fixed elasticity model (model IVA)

Variable	Estimate	<i>t</i> value
λ_1	−0.4875	−53.56
λ_2	−1.2193	−69.79
β_1 (log Income)	0.5292	54.18
β_2 (Hat2000)	−0.0357	−10.25
β_3 (Hat2001)	−0.0008	−0.22
β_4 (Hat2002)	0.0106	2.95
Observations	48,015	
$\hat{\sigma}^2$	0.034676	
Log likelihood	14,235.26	
AIC	−22,054.5	

figures contained in the Passenger Demand Forecasting Handbook (PDFH) that contains the elasticities recommended for use in the railway industry in Great Britain (ATOC 2013). The GJT elasticity is very similar to PDFH's value of -1.2 , although the PDFH figure is explicitly long run. The recommended PDFH income elasticity is 1.4 between major cities and 0.65 for other flows. The income elasticity estimated here is smaller than the values that PDFH recommends, although not greatly so given the other flows dominate our data. The fare elasticity is lower than PDFH's -1.2 , but we should note that PDFH is intended to give long run fare elasticities and the static model of Table 4 will understate the long run effect.

The statistics presented in this table (other than the parameter estimates and the associated t ratios are calculated as follows. The sum of squares of errors is defined as usual in regression analysis by $SSE = (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}})$, where the observed and predicted vectors \mathbf{y} and $\hat{\mathbf{y}}$ are defined by $y = \log T$, with T indicating the number of trips. Then,

- the mean squared error $\hat{\sigma}^2 = SSE/(R - h)$, where R is the number of observations and h is the number of estimated parameters;
- the log likelihood is calculated by (Spiess and Neumeyer, 2010)⁵ $\log(L) = 0.5(-R(\log 2\pi + 1 - \log R + \log SSE))$;
- the Akaike Information Coefficient (AIC) is calculated by

$$AIC = 2(h + 1) - 2\log(L)$$

The positive values for log likelihood are unusual, but arise from the calculation based on the assumption that the error term is normally distributed. All three of these statistics are measures of model quality and we can see from the formulae that for a good model the mean squared error and AIC should be small (less positive and more negative respectively), while the log likelihood should be large (more positive).

Undamped model

In contrast, the model suggested by simple linear functions of time and cost, as might be suggested by conventional urban transport planning, i.e. Model 0 discussed in Sect. 2.3, is presented in Table 5.

Comparing the models of Tables 4 and 5, it is clear that the undamped model gives a substantially worse explanation of the data than the classical rail model and can be rejected. That is, the simplest transfer of a linear model is much worse for rail ticket data than the models used by rail data analysts.

Distance-damped models

The third group of models considered are drawn from conventional transport planning with distance damping, i.e. models of Type I and II in the discussion of Sect. 2.3. For these models, the function f appears as

$$\text{Type I: } f = (\lambda_1 \text{fare} + \lambda_2 \text{GJT}) \cdot \text{dist}^{-\alpha}$$

$$\text{Type II: } f = \lambda_1 \text{fare} \cdot \text{dist}^{-\alpha_1} + \lambda_2 \text{GJT} \cdot \text{dist}^{-\alpha_2}$$

⁵ We report log likelihood rather than the sum of squares of errors because this permits the use of χ^2 tests for the significance of the difference of nested models.

Table 5 Undamped model (model 0)

Variable	Estimate	<i>t</i> value
λ_1	−0.0045	−12.48
λ_2	−0.0034	−44.81
β_1 (log Income)	0.8832	98.94
β_2 (Hat2000)	−0.0206	−5.64
β_3 (Hat2001)	0.0113	3.05
β_4 (Hat2002)	0.0164	4.37
Observations	48,015	
$\hat{\sigma}^2$	0.038297	
Log likelihood	11,850.77	
AIC	−17,285.5	

The results of these analyses are given in Table 6.

The fact that two of the α values are significantly outside the acceptable range (0, 1) to obtain cost damping and pass the kilometrage test may indicate that these formulations are simply not valid. However, the indication is that the α values are strongly positive and that (in Type II) fare damping is stronger than time damping, so that the value of time increases with trip length, as expected.

In addition to the unacceptable α values, the fit offered by these models is worse than the standard rail model, so these models do not offer any useable advance.

‘Dynamic’ functional forms

The final group of models estimated were those in which the damping operates as a function either of the generalised cost or of its components. Thus the damping is an intrinsic, ‘dynamic’ part of the model, rather than as an additional component as in the distance damping.

In models III and IIIA, the damping operates on the entire utility function. For these models, the function f appears as

$$\text{Type III: } f = \lambda_0 (\text{fare} + \lambda_2 GJT)^\alpha$$

$$\text{Type IIIA: } f = \lambda_0 \log(\text{fare} + \lambda_2 GJT)$$

Model III proved difficult to estimate, as the combination of λ_0 and α introduced too much correlation. Tests were made fixing α and the best results (reported below) were obtained by setting $\alpha = 0.08$. It seemed that α values even closer to zero might give still better results, but these presented too many numerical problems. An attempt was made to estimate the Box-Cox model, which generalises both III and IIIA

$$\text{Type III Box-Cox: } f = \lambda_0 \frac{(\text{fare} + \lambda_2 GJT)^\alpha - 1}{\alpha}$$

but this also required fixing λ_0 (at −1.0) and did not improve on the $\alpha = 0.08$ result, but gave a best-estimate $\alpha = 0.083$, suggesting that the value of α should indeed be around 0.08.

The results of these analyses are given in Table 7.

Table 6 Distance-damped models

Variable	Type I		Type II	
	Estimate	<i>t</i> value	Estimate	<i>t</i> value
λ_1	−4.7632	−14.73	−35.70	−12.52
λ_2	−0.7292	−17.12	−0.1132	−9.69
α	1.1111	79.22	n.a.	
α_1	n.a.		1.5424	79.33
α_2	n.a.		0.6690	29.86
β_1 (log Income)	0.6835	73.00	0.6299	66.68
β_2 (Hat2000)	−0.0349	−9.83	−0.0322	−9.17
β_3 (Hat2001)	−0.0027	−0.74	0.0015	0.42
β_4 (Hat2002)	0.0046	1.25	0.0094	2.61
Observations	48,015		48,015	
$\hat{\sigma}^2$	0.035993		0.035275	
Log likelihood	13,340.93		13,825.40	
AIC	−20,263.9		−21,230.8	

Table 7 Dynamically-damped models (entire function)

Variable	Type III ^a		Type IIIA	
	Estimate	<i>t</i> value	Estimate	<i>t</i> value
λ_0	−9.3867	−74.66	−1.5352	−76.11
λ_2	24.5461	35.22	0.2231	36.66
α	0.08	Fixed	n.a.	
β_1 (log Income)	0.6364	67.08	0.6216	65.52
β_2 (Hat2000)	−0.0327	−9.26	−0.0337	−9.57
β_3 (Hat2001)	0.0024	0.68	0.0011	0.32
β_4 (Hat2002)	0.0124	3.41	0.0112	3.08
Observations	48,015		48,015	
$\hat{\sigma}^2$	0.035690		0.035508	
Log likelihood	13,543.67		13,666.50	
AIC	−20,671.3		−20,917.0	

^a In this model, the fare is represented in pence, not pounds, to facilitate estimation

These results indicate that the log function of IIIA outperforms the power functions of Type III. In fact, the log function is the limiting case of the Box–Cox when $\alpha \rightarrow 0$, so that the result and the difficulty in estimating positive values of α are not entirely surprising.

It can be seen that the model IIIA gives a fit to the data between those of Types I and II but not as good as that of the classical rail model reported in Sect. 3.2, which is very similar in form.

In the final group of models, Type IV, the damping operates on the separate time and cost variables. The log form of this model corresponds to the conventional rail models and is reported in Sect. 3.2. A further model form investigated here was

$$\text{Type IV: } f = \lambda_1 \cdot \text{fare}^{\alpha_1} + \lambda_2 \cdot GJT^{\alpha_2}$$

Model IV also did not converge, so that the coefficients λ had to be set to -1.0, and the Box-Cox variant also failed in the same way; however, the Box-Cox variant gave for the first time a model fit (with negative α) better than the fixed-elasticity model IVA (which is the limit of the model as $\alpha \rightarrow 0$). Fixing α gave robust estimation results for λ , with the value 0.03 for both α parameters being the best that could be obtained. It seemed that smaller values would give still better results but again numerical problems prevented us from obtaining results. The results for the acceptable model forms are shown in Table 8.

The final model of this family that was tested was Model IVB, which represents a different approach to transforming time and cost.

$$\text{Type IVB: } f = \left(\begin{array}{l} \alpha_1 \lambda_1 \text{fare} + q_1 (1 - \alpha_1) \lambda_1 \log(\text{fare}) \\ + \alpha_2 \lambda_2 \text{GJT} + q_2 (1 - \alpha_2) \lambda_2 \log(\text{GJT}) \end{array} \right)$$

where $q_1 = \frac{\text{fare}}{\log(\text{fare})}$ and $q_2 = \frac{\text{GJT}}{\log(\text{GJT})}$

Rather than the Box-Cox or power function, this model gives a flexible form by combining log and linear forms of each variable, mixed using linear factors. While not linear in its parameters (α multiplies λ), this form imposes much less burden on the estimation procedure while giving the same freedom as a Box-Cox transformation. The parameters α give an indication of the curvature of the function and can be interpreted analogously to the exponent of Box-Cox or power functions. The results of this model are shown in Table 9.

Here we obtain a fit to the data which is substantially better than that of the classical rail models. That is, the cost damping given by the log function, in principle the strongest damping that can be applied in a smooth monotonic function, appears inadequate. However, the model presents an issue because of the sign of both α_1 and α_2 , which imply that for sufficiently large values of GJT or cost the slope of the function will be incorrect, i.e. that increases in time or cost would imply an increase in demand.

This issue can be investigated by calculating the derivatives of f with respect to GJT and cost. We obtain

Table 8 Dynamically damped models (power functions of components)

Variable	Type IV		Type IV fixed α	
	Estimate	<i>t</i> value	Estimate	<i>t</i> value
λ_1	-1.0	Fixed	-14.6403	-52.47
λ_2	-1.0	Fixed	-34.7155	-69.51
α_1	0.2006	64.29	0.03	Fixed
α_2	0.2705	155.84	0.03	Fixed
β_1 (log Income)	0.6340	66.10	0.5360	54.83
β_2 (Hat2000)	-0.0313	-8.87	-0.0355	-10.18
β_3 (Hat2001)	0.0037	1.03	-0.0004	-0.12
β_4 (Hat2002)	0.0137	3.78	0.0109	3.04
Observations	48,015		48,015	
σ^2	0.035563		0.034773	
Log likelihood	13,629.10		14,168.40	
AIC	-20,842.2		-21,920.8	

Table 9 Dynamically damped models (mixed log and linear functions)

Variable	Type IVB	
	Estimate	<i>t</i> value
λ_1	−0.1073	−63.55
λ_2	−0.0289	−47.85
α_1	−0.1801	−44.69
α_2	−0.0216	−5.03
β_1 (log Income)	0.5109	52.74
β_2 (Hat2000)	−0.0331	−9.63
β_3 (Hat2001)	−0.0012	−0.35
β_4 (Hat2002)	−0.0091	2.57
Observations	48,015	
$\hat{\sigma}^2$	0.033655	
Log likelihood	14,954.25	
AIC	−23,488.5	

$$\frac{\partial \log T}{\partial t} = \lambda_2 \left(\alpha_2 + \frac{q_2(1 - \alpha_2)}{GJT} \right)$$

This is positive when $GJT > 2139$ min (over 35 h, one way) and we have no observations of this magnitude in our data set, which was limited to 300 miles (one way). Similarly,

$$\frac{\partial \log T}{\partial c} = \lambda_1 \left(\alpha_1 + \frac{q_1(1 - \alpha_1)}{c} \right)$$

becomes positive when the adjusted fare $c > 46.03$ (pounds), which is the case for 2.1 % of the data that was analysed. This means that there is a real limitation to the use of model IVB, but its improved fit to the data means that the model offers new insight into the behaviour of train passengers.

Discussion and conclusions

The objective of this study was to investigate whether insights and functional forms derived from conventional (road based) transport planning could be applied to the analysis of rail ticket sales data. This data has typically been analysed using constant-elasticity models which, viewed from the standpoint of conventional transport planning, imply an extreme form of cost damping, i.e. that sensitivity to cost (and time) decreases as journey length increases. Cost damping has been observed in conventional planning work, but perhaps not to the same extent as in rail data analysis.

Initial investigation suggested that it was possible, at least as an approximation, to interpret the functional forms of conventional transport planning for use in rail data analysis. Functions were therefore set up that mirrored the four forms of cost damping identified in the work by Daly (2010), as well as a model without cost damping. It turned out that one of these four forms reproduced the fixed-elasticity model of classical rail data analysis.

The first result of the analysis is that the existence of cost damping is strongly confirmed by this data. Very low values of exponents are found, close to zero, so that log functions are suggested. The second result is that damping of cost is stronger than damping of time, so that the value of time increases with trip length. Both of these results confirm findings from other models. They strengthen the case for including damping, if possible differentially by time and cost, in travel demand models generally. Omission of this effect could cause errors in forecasting, depending on the nature of the forecast scenarios being considered.

Estimating income elasticity and the impact of the Hatfield accident were not among the objectives of the study, but the values that were obtained in the better models were plausible.

The detailed results of the models are summarised in Table 10. The fit of each model is indicated by its log likelihood; use of the AIC measure would not change the order of preference of the models. The elasticities are calculated for mean values of time, cost and distance as observed in the data used for analysis; values for the sample population could also be calculated but this was beyond the scope of the current work.

Here we see that models of Type IV, with component-specific ‘dynamic’ damping, fit the data better than the other models. Of the other models, Model 0 fits very poorly and has unacceptable cost elasticity and value of time results. Models I and II fail the kilometrage test. Model III requires α to be fixed but it and Model IIIA are primarily rejected for models of Type IV because of their inferior fit.

Within the models of Type IV, the base model IV is less good, because it requires us to fix α , it fits the data less well and the implied value of time is a little high. There is little to choose between models IVA and IVB in terms of the elasticity or implied value of time. Following the discussion of Sect. 3.2, the time elasticity value appears reasonable, comparing it with the PDFH equivalent; while these models are less elastic to cost than PDFH recommends, they are well within the broad range suggested by the UK Department for Transport (2015), which is -0.9 to -0.2 . The value of time is also slightly higher than we might expect, suggesting that the effect of fares is not fully captured in these models. For this reason, and because travellers can switch between ticket types to mitigate the effect of fare increases, we do not recommend analysis of this type for the estimation of values of time.

In summary, this work has shown that it is possible to apply the functional forms used in conventional transport demand analysis to rail ticket sales data. When this is done, a high

Table 10 Output of proposed model forms

Type	Fit (log-likelihood)	Range issues	Cost elasticity (at mean)	Time elasticity (at mean)	Value of time (p /minute, at mean)
0	11,851	No	-0.09	-0.82	75.6
I	13,341	$\alpha > 1$	-0.39	-0.76	15.3
II	13,825	$\alpha_1 > 1$	-0.35	-1.03	23.0
III	13,544	α fixed	-0.37	-1.16	24.5
IIIA	13,667	No	-0.40	-1.13	22.3
IV	14,168	α fixed	-0.36	-1.19	26.0
IVA	14,235	No	-0.49	-1.22	19.7
IVB	14,954	$\alpha < 0$	-0.52	-1.19	17.9

degree of cost damping is found, stronger than given by a log function. Further analyses should look at the methods available for representing this extreme form of damping.

Acknowledgments Nobuhiro Sanko acknowledges support by JSPS KAKENHI Grant Number 24330132. An earlier version of this paper was presented at the hEART Conference, Leeds in September 2014. Comments by anonymous reviewers and the editor have helped us improve the paper, but we remain responsible for interpretations and any errors.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Abrantes, P.A.L., Wardman, M.: Meta-analysis of UK values of time: an update. *Transp Res A* **45**(1), 1–17 (2011)
- Aptech Systems, Inc. GAUSS. <http://www.aptech.com/> (2015). Accessed 13 Nov 2015
- Association of Train Operating Companies—ATOC: Passenger Demand Forecasting Handbook, Version 5.1. ATOC, London (2013)
- Daly, A.: Cost Damping in Travel Demand Models, Report for Department for Transport. <http://webarchive.nationalarchives.gov.uk/20110202223908/>, <http://www.dft.gov.uk/pgr/economics/rdg/costdamping/> (2010)
- Daly, A., Carrasco, J.: The influence of trip length on marginal time and money values. In: Kitamura, R. et al. (eds.) *The Expanding Sphere of Travel Behaviour Research: Selected Papers from the Proceedings of the 11th Conference on Travel Behaviour Research*. Emerald Books (2009)
- Daly, A., Fox, J.: Forecasting mode and destination choice responses to income change. In: Presented at IATBR, Toronto (2012)
- Department for Transport: (2014) TAG Unit M2, Section 3.3 on Cost Damping: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/275597/webtag-tag-unit-m2-variable-demand-modelling.pdf#nameddest=chptr03
- Department for Transport. TAG Unit M2, Section 6.4 on Realism Testing. <https://www.gov.uk/government/publications/webtag-tag-unit-m2-variable-demand-modelling> (2015)
- Fox, J., Daly, A., Gunn, H.: Review of RAND Europe's Transport Demand Model Systems, RAND Europe Report MR 1694. http://www.rand.org/pubs/monograph_reports/MR1694.html (2003)
- Hill, C., Adkins, L.: Using Gauss for Econometrics. <http://pages.suddenlink.net/ladkins/pdf/GAUSS.pdf> (2001). Accessed 4 Aug 2013
- Jones, I.S., Nichols, A.J.: The demand for inter-city rail travel in the United Kingdom: some evidence. *J. Transp. Econ. Policy* **17**(2), 133–153 (1983)
- Rich, J., Mabit, S.L.: Box–Cox approximations and beyond. In: Presented to hEART Conference, Stockholm (2013)
- Spieß, A.-N., Neumeyer, N.: An evaluation of R^2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach. *BMC Pharmacol.* (2010). doi:10.1186/1471-2210-10-6
- Wardman, M.: Price elasticities of surface travel demand: a meta-analysis of UK evidence. *J. Transp. Econ. Policy* **48**(3), 367–384 (2014)

Andrew Daly is a Research Professor at the Institute for Transport Studies at the University of Leeds, Visiting Professor at the Royal Institute of Technology in Stockholm, Senior Research Fellow at RAND Europe in Cambridge and the author of the widely-used ALOGIT software. He received the 2012 Lifetime Achievement Award from the International Association for Travel Behaviour Research.

Nobuhiro Sanko is an Associate Professor at the Graduate School of Business Administration, Kobe University. He holds Doctor of Engineering and Master of Engineering degrees from Nagoya University and an MBA from Ecole Nationale des Ponts et Chaussées. His research interests include the long-term

dynamics of travel behaviour, stated preference, and transport planning. A part of this work was conducted when he was a visiting research fellow at the Institute for Transport Studies, University of Leeds.

Mark Wardman was until recently Professor of Transport Demand Analysis at the Institute for Transport Studies, University of Leeds and retains a visiting position there. He is currently Technical Director at SYSTRA Ltd. Mark has researched travel behaviour for over 30 years, with a particular interest in the demand for rail travel upon which subject he has published widely. He has conducted extensive reviews of rail price elasticities and time elasticities, and valuations of time, reliability, crowding and improved rolling stock.